UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/782,434	02/13/2001	Martin Franz	YOR9-2001-0011US1 (8728-4	9870

46069 7590 04/08/2005

F. CHAU & ASSOCIATES, LLC
130 WOODBURY ROAD
WOODBURY, NY 11797

EXAMINER	
JACKSON, JAKIEDA R	
ART UNIT	**PAPER NUMBER**
2655	

DATE MAILED: 04/08/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
P.O. Box 1450
ALEXANDRIA, VA 22313-1450
www.uspto.gov

MAILED
APR 08 2005
Technology Center 2600

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 09/782,434
Filing Date: February 13, 2001
Appellant(s): FRANZ ET AL.

Koon Hon Wong
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed January 5, 2005.

(1) Real Party in Interest

A statement identifying the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The brief contain a statement identifying that there is no related appeals and interferences which will directly affect or be directly affected by or have a bearing on the decision in the pending appeal is contained in the brief.

(3) Status of Claims

The statement of the status of the claims contained in the brief is correct.

(4) Status of Amendments After Final

The appellant's statement of the status of amendments after final rejection contained in the brief is correct.

(5) Summary of Invention

The summary of invention contained in the brief is correct.

(6) Issues

The appellant's statement of the issues in the brief is correct.

(7) Grouping of Claims

The rejection of claims 1 and 3-24 stand or fall together because appellant's brief does not include a statement that this grouping of claims does not stand or fall together and reasons in support thereof. See 37 CFR 1.192(c)(7).

(8) Claims Appealed

The copy of the appealed claims contained in the Appendix to the brief is correct.

(9) Prior Art of Record

6,167,377	Gillick et al.	12-2000
5,625,749	Goldenthal et al.	04-1997
5,621,809	Bellegarda et al.	04-1997
5,950,158	Wang	09-1999

(10) Grounds of Rejections

The following ground(s) of rejection are applicable to the appealed claims:

Claims 1 and 3, 5-13, 15-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Gillick et al, hereinafter referenced as Gillick.

Regarding **claim 1, 12, 20 and 21**, Gillick discloses an Automatic Speech Recognition (ASR) system (figure 1, element 160; column 1, lines 6-7) having at least two language models (variety of language models; column 2, lines 1-5), a method for combining language model scores (column 16, lines 8-11) generated by at least two language models, said method comprising the steps of:

generating a list (figure 11, element 1125) of most likely words for a current word in a word sequence uttered by a speaker (column 1, lines 33-42), and acoustic scores corresponding to the most likely words (figure 9);

computing language model scores for each of the most likely words in the list (column 18, lines 36-39), for each of the at least two language models;

respectively and dynamically determining a set of coefficients (column 16, lines 20-40) to be used to combine the language model scores of each of the most likely

words in the list (column 16, lines 8-11) based on a context of the current word (column 17, lines 39-41);

respectively combining the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the set of coefficients determined therefor (column 16, lines 8-11);

wherein said determining step comprises the steps of:

dividing text data (column 1, lines 8-13) for training (column 15, lines 7-13) a plurality of sets of coefficients into partitions (frames; column 1, lines 8-13), depending on words counts (identifying words/scores) corresponding to each of the at least two language model (utterance/language models; column 15, lines 60-67 with column 16, lines 20-32 and lines 44-48); and

for each of the most likely words in the list, dynamically selecting (figure 4A, element 405) the set of coefficients from among the plurality of sets of coefficients so as to maximize the likelihood (likelihood of the match; column 4, lines 1-20) of the text data with respect to the at least two language models (column 4, lines 46-67).

Regarding **claims 3, 13 and 22**, Gillick discloses the method wherein the at least two language models comprises a first and second language model, and said dividing step comprises the step of grouping, in a same partition, word triplets sub.1w.sub.2w.sub.3 (trigram models) which have a count for the word pair w.sub.1w.sub.2 (bigram models; column 1, line 63 – column 2, line 5 and pair of words; column 14, lines 17-32) in first language model (first, second and/or third language

models) greater than the count for the word pair w.sub.1w.sub.2 in the second language model (fourth language model; column 18, lines 16-28).

Regarding **claims 5 and 15**, Gillick discloses the method further comprising the step of, for each of the most likely words in the list, combining an acoustic score (acoustic score) and the composite score (previous score) to identify a group of most likely words to be further processed (column 10, lines 8-14).

Regarding **claims 6, 16 and 23**, Gillick discloses the method wherein the group of most likely words contains less words than the list of most likely words (added to the list of words; column 7, lines 60 – column 8, lines 32).

Regarding **claim 7**, Gillick discloses the method wherein the partitions are independent from the at least two language models (column 2, lines 1-5).

Regarding **claim 8**, Gillick discloses the method further comprising the step of representing the set of coefficients by a weight vector comprising n-weights (interpolation weights), where n (lambda 1 and 2) equals a number of language models in the system (column 16, lines 1-25), to identify the best corresponds to a user's utterance.

Regarding **claims 9, 17 and 24**, Gillick discloses the method wherein said combining step comprises the steps of:

for each of the most likely words in the list (column 1, lines 43-47),

multiplying a coefficient corresponding to a language model by a language model score corresponding to the language model to obtain a product for each of the at least two language models (column 10, lines 16-18); and

summing the product for each of the at least two language models (column 10, lines 8-67), in order to determine the acoustic models that best matches the utterance.

Regarding **claims 10 and 18**, Gillick discloses the method wherein the text data for training the plurality of sets of coefficients is different than language model text data used to train the at least two language models (column 16, lines 26-29).

Regarding **claim 11**, Gillick discloses a method for combining language model scores (column 16, lines 8-11) generated by at least two language models (variety of language models; column 2, lines 1-5) comprised in an Automatic Speech Recognition (ASR) system (figure 1, element 160; column 1, lines 6-7), said method comprising the steps of:

generating a list (figure 11, element 1125) of most likely words for a current word in a word sequence uttered by a speaker (column 1, lines 33-42), and acoustic scores corresponding to the most likely words (figure 9);

computing language model scores for each of the most likely words in the list (column 18, lines 36-39), for each of the at least two language models;

respectively and dynamically determining a weight vector to be used to combine the language model scores of each of the most likely words in the list based on the context of the current word (column 16, lines 8-11 with column 17, lines 39-41), the weight vector comprising n-weights (interpolation weights), wherein n (lambda 1 and 2) equals a number of language models in the system (column 16, lines 1-25), and each of the n-weights depend upon n-gram history counts (frequency of words; column 14, lines 26-32 with column 16, lines 44-48); and

respectively combining the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the set of coefficients determined therefor (column 16, lines 8-11).

Regarding **claim 19**, Gillick discloses a combining system for combining language model scores (column 16, lines 8-11) generated by at least two language models (variety of language models; column 2, lines 1-5) comprised in an Automatic Speech Recognition (ASR) system (figure 1, element 160; column 1, lines 6-7), the ASR system having a fast match (processor) for generating a list (figure 11, element 1125) of most likely words for a current word in a word sequence uttered by a speaker and acoustic scores corresponding to the most likely words (column 1, lines 33-42) combing system comprising:

a language model score computation device (hardware or software) adapted to compute language model scores for each of the most likely words in the list (column 18, lines 36-39), for each of the at least two language models;

a selection device (recognizer; figure 13, element 215) adapted to respectively and dynamically select a weight vector to be used to combine the language model scores of each of the most likely words in the list based on the context of the current word (column 16, lines 8-11 with column 17, lines 39-41), the weight vector comprising n-weights (interpolation weights), wherein n (lambda 1 and 2) equals a number of language models in the system (column 16, lines 1-25), and each of the n-weights depend upon n-gram history counts (frequency of words; column 14, lines 26-32 with column 16, lines 44-48); and

a combination device (select command; column 4, lines 46-50) adapted to respectively combining the language model scores of each of the most likely words in the list to obtain a composite score for each of the most likely words in the list, using the set of coefficients determined therefor (column 16, lines 8-11).

Claims 4 and 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gillick in view of Goldenthal et al., hereinafter referenced as Goldenthal.

Regarding **claims 4 and 14**, Gillick discloses speech recognition language models, but lacks disclosing the method wherein said selecting step comprises the step of applying the Baum Welch iterative algorithm to the plurality of sets of coefficients.

Goldenthal discloses the method wherein said selecting step comprises the step of applying the Baum Welch iterative algorithm to the plurality of sets of coefficients (column 2, lines 41-43), for training Hidden Markov Models (HMM's).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Gillick's invention such that it applied the Baum Welch iterative algorithm, in order to handle speech problems (column 2, lines 31-32).

(11) Response to Arguments

Appellants asserts on pages 5-7 of the appeal brief, that Gillick fails to describe "determining a set of coefficients to be used to combine the language model scores, based on a context of the current word," as claimed in claim 1. However, the Examiner maintains that because Gillick does disclose ***determining a set of coefficients*** (lambda 1 and lambda 2; column 16, lines 20-40) to be used to combine the language

model scores (combine the scores produced by the language models; column 16, lines 8-11) of each of the most likely words in the list (assigns weights; column 15, lines 60-67), ***based on a context of current word*** (column 16, lines 50-59). The equations involved teach that the technique is updated based on what happened previously, which is context.

Appellants further assert on pages 7 and 9-10 of the appeal brief that Gillick fails to describe "determining a weight vector to be used to combine the language model scores of each of the most likely words in the list based on a context of the current word, the weight vector comprising n-weights, wherein n-equals a number of language models in the system, and each of the n-weights depends upon history n-gram count", as claimed in claims 11 and 19. However, the Examiner maintains that because in addition to what is previously mentioned above, in regards to claim 1, Gillick further discloses "***determining the weight vector*** to be used to combine the language model scores of each of the most likely words in the list (lambda 1 and lambda 2; column 16, lines 20-40), the weight vector comprising n-weights, wherein n equals a number of language models in the system (column 16, lines 20-21), and each of the ***n-weights depends upon history n-gram counts*** (equation; column 16, line 15 is calculated in the past) meanwhile, (equation; column 16, lines 50-59 is calculated presently).

Appellants assert on pages 8-9 of the appeal brief, that Gillick fails to describe "dividing text data for training a plurality of sets of coefficients into partition, depending on word counts corresponding to each of the at least two language modes," as claimed in claim 1. However, the examiner maintains that Gillick discloses ***dividing*** (takes a

part of k words, not all) ***text* data** (recognition word is text for $w_1, w_2 \ldots w_k$) for training a plurality of sets of coefficients into partitions depending on word counts corresponding to each of the at least two language models (to identify the best recognition candidates column 16, lines 44-48 with lines 20-32).

Also, appellants assert on page 8 of the appeal brief that the examiner admitted that in paper no. 4, page 2 Gillick does not discloses "dividing text data". However, after carefully reviewing paper no. 4, page 2, it states:

> Applicant argues, regarding claim 2, that Gillick does not disclose "dividing text data", instead Gillick discloses dividing the spoken utterance. Applicant also argues, regarding claim 2, that Gillick does not anticipate "dividing text data for training a plurality of sets of coefficients into partitions, depending on word counts corresponding to each of the at least two language models".

Nowhere in that passage does it shows that the Examiner admits that Gillick does not disclose dividing text data. In fact, Gillick teaches dividing text, as explained above.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

JRJ
April 4, 2005

DAVID L. OMETZ
PRIMARY EXAMINER

Conferees
David Ometz
Richemond Dorvil
Tālivaldis Šmits

Frank Chau, Esq.
F. CHAU & ASSOCIATES, LLP
Suite 501
1900 Hempstead Turnpike
East Meadow, NY 11554